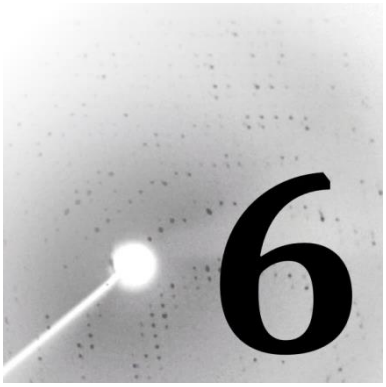
Just add sugar...

Pijning, Tjaard



Summary & Outlook

In this thesis, we have seen how expanding our knowledge of the structural aspects of glucan- and fructansucrases from *Lactobacilli* contributes to a better understanding of their workings. Still, many questions remain for these intriguing enzymes. In this chapter we will summarize our findings, and give suggestions for future research that ultimately should provide us with the tools to use the enzymes and their products in practical applications.

Carbohydrates constitute one of the major classes of biomolecules, and are found in all kingdoms of life. They perform a wide variety of functions, such as energy storage, communication, provision of structural integrity, and defense and protection against hostile environments. An example of the latter function is given by the extracellular polysaccharides (EPS) produced by certain lactic acid bacteria (LAB) and other bacterial species. Because of its sticky properties, the extracellular carbohydrate layer enables the bacteria to form biofilms, and adhere to surfaces. This happens for example in the human oral cavity, where EPS-producing *Streptococcus* species may give rise to dental plaque and tooth decay. Biofilm formation may also be beneficial: LAB colonize the human gastrointestinal tract, and help digest nutrients that cannot be processed by gastric enzymes. It appears that LAB are beneficial for the intestinal microbial ecosystem because they suppress the growth of pathogens (the so-called probiotic effect). In turn, food-borne carbohydrates that pass the stomach unprocessed and are digested by probiotic bacteria in the gastrointestinal tract, have a similar effect, and are called prebiotics. Because of these beneficial health effects, LAB and the EPS they produce have gained substantial interest for food-related applications. Moreover, EPS have interesting properties for non-food and medical applications. A large number of EPS-producing LAB strains have been identified in recent years, as well as the specific carbohydrates they produce.

Extracellular polysaccharides synthesized by LAB show an enormous structural variation, leading to a range of physicochemical properties such as solubility, viscosity, water-binding ability and stickiness. Structure-wise, EPS of the α-glucan type have a α-D-glucose as the building block and may contain up to ~250,000 sugar units; β-fructan type EPS may contain up to ~1,000,000 β-D-fructose units. Recently derived models show that α-glucans have a complex but well-defined and specific structure (Leemhuis *et al.*, 2013a); β-fructans in general are less complex (Vijn & Smeekens 1999).

In EPS-producing bacterial strains, the polysaccharides are synthesized by specialized 'machineries', extracellular enzymes called glucansucrases (GSs) and fructansucrases (FSs). To do this, they only require a simple and cheap substrate, sucrose. Understanding how the different GSs and FSs synthesize their specific products would allow us to use them to our advantage, *e.g.* by engineering the enzymes such that they produce carbohydrates with desired properties. In trying to relate structure to specific function, an important tool is to determine the structures of these intriguing machineries. This thesis describes the structural insights that we have obtained from several GSs and FSs from LAB.

GLUCANSUCRASES

At present, ~230 putative bacterial GSs are listed in the CAZy database, but only 57 of these have been characterized, displaying GS activity. Bacterial GSs are large enzymes (120-200 kDa), and are classified in glycoside hydrolase family 70 (GH70). They are closely related to carbohydrate-processing enzymes from GH13 and GH77, sharing evolutionary, structural and mechanistic similarities.

GSs from lactic acid bacteria (LAB) were identified first by Van Geel-Schutten *et al.* (1999). Before structural information was available for GH70 GSs, it was proposed that they would contain a catalytic domain harboring a $(\beta/\alpha)_8$ barrel, flanked by variable N- and C-terminal domains in a 'linear' domain organization. Biochemical and structural investigation of LAB GSs began about a decade ago (Kralj *et al.*, 2002; Kralj *et al.,* 2004a,b), when recombinant GS constructs from different *Lactobacillus reuteri* species could easily be overexpressed and purified. Structural work was then initiated; the N-terminally truncated glucansucrase GTFA-ΔN from *L. reuteri* 121 was the first GS to be crystallized (~2003). However, due to the poor diffraction properties of the GTFA-ΔN crystals, the structure could only be determined much later. However, soon after crystallization of GTFA-ΔN, a close homologue, GTF180-ΔN from *L. reuteri* 180 was also successfully crystallized, as described in **Chapter 2**. The construct used for crystallization lacked the N-terminal ~700-residue part, but was fully active and had similar biochemical and kinetic properties as the full-length enzyme. Three different well-diffracting crystal forms of GTF180-ΔN were obtained, allowing determination of the first GH70 GS structure (Vujičić-Žagar *et al.*, 2010). One of the most striking features of this structure was the domain arrangement: instead of the proposed linear arrangement, a rare 'U-shape' fold was observed in which four of the five domains are constituted from *two* polypeptide segments (Chapter 1: Figure 4b,c). Moreover, apart from three domains (A, B and C) also present in GH13 enzymes, two additional domains (IV and V) with so far unknown function were found. The structure determination of two more GSs, GTF-SI from *Streptococcus mutans* (Ito *et al.*, 2011) and DSR-E ΔN_{123}-GBD-CD2 (Brison *et al.*, 2012) from *Leuconostoc mesenteroides* NRRL B-1299 followed soon after, confirming the common domain organization of GH70 GSs. Moreover, GS complexes with substrate (sucrose), acceptor (maltose) or inhibitor (acarbose) confirmed the proposed structural details of the catalytic mechanism, and provided insight into the way GSs perform their function (see below). The fourth GS crystal structure, GTFA-ΔN from *L. reuteri 121*, described in **Chapter 4**, was determined from crystals that diffracted relatively poorly, perhaps related to the fact that they had an exceptionally high solvent content and large solvent channels. The overall structure of GTFA-ΔN resembles that of GTF180-ΔN, further confirming the GS domain organization and shared substrate

specificity. Moreover, the GTFA-ΔN structure supports previous ideas about product specificity (see below). Thus, even low-resolution GS structures hold important information with respect to their function and specificity.
All four published GS structures are of truncated enzymes; the 'missing' N-terminal parts may hold important information concerning GS function as well, but structures of full-length GSs have not been reported yet. In our laboratory we have taken different approaches aiming at this. First, in **Chapter 3** we describe for the first time the molecular shape of a full-length glucansucrase (GTF180), obtained using small-angle X-ray scattering (SAXS). The ~700 N-terminal residues form a long extended domain and appear to constitute a well-folded, non-flexible entity. This should be a favorable property with regard to successful crystallization. Indeed, in another approach, full-length GTF180 and GTFA have been successfully crystallized in our laboratory (unpublished results); however, despite extensive efforts to optimize crystallization conditions and cryo-cooling methods, the diffraction properties of the crystals were insufficient to allow structure determination. Therefore, other strategies need to be applied, such as variation of the constructs used for crystallization (*e.g.* omitting small segments from the N- or C-terminus). Another strategy could be to crystallize a construct containing only the N-terminal ~700 residues.

From the GS structures, the functions of domains A and B of GSs are obvious. Domain A harbors the catalytic site, with the 3 catalytic residues near the bottom of a deep pocket, where the substrate binds. Loops from domain A and B flank a rather wide groove at the surface, which probably forms the binding site for (intermediate) products.
Regarding domain V, the GS structures showed that the sequence repeats found in this domain form a β-solenoid fold that is also observed in proteins that bind carbohydrate moieties present in cell wall components. This strengthens the hypothesis that the N- and C-terminal sequence repeats in GSs are involved in binding α-glucans and/or cell wall associated carbohydrate components. However, soaking of GS crystals with different (small) sugars did not reveal any binding in domain V. Perhaps only longer carbohydrate chains can bind (*e.g.* from a growing α-glucan chain); another possibility is that additional binding sites in the N-terminal region are necessary to bind such carbohydrate chains.
Interestingly, a comparison of the four GS structures revealed that domain V can adopt different orientations with respect to the other domains. Even more so, in a second crystal form of GTF180-ΔN, described in **Chapter 3**, domain V is rotated by ~120°, such that the enzyme adopts a much more compact conformation. To investigate whether the intrinsic flexibility would be functional during α-glucan synthesis (*e.g.* resulting in a 'swinging' motion of α-glucan binding domains during chain elongation), we determined the solution structures of GTF180-ΔN and GTF180 (**Chapter 3**). Both the truncated and

the full-length enzyme adopt a more or less elongated conformation with no indications of flexible parts, even under conditions where the enzyme synthesizes small oligosaccharides. We conclude that the intrinsic flexibility does not seem to play a role during early stages of the GS reaction. It may be that flexibility becomes important only when longer oligosaccharides/polymer are being formed; however technical limitations prevent such conditions to be tested by SAXS. Another approach would be to analyze reaction kinetics of a construct lacking intrinsic flexibility, *e.g.* engineering disulfide bridge formation between domains V and B by a double Cys mutation. Such experiments have been initiated but need to be worked out further (S. Kralj, personal communication).

Structures of GSs and their complexes with substrate, acceptor or inhibitor confirmed that they have a single active site and use the same catalytic mechanism as GH13 enzymes. In contrast to GH13 enzymes, the active site of GH70 GSs is blocked beyond subsite -1, explaining why GH70 glucansucrases only transfer a single glucosyl unit in each reaction cycle. In the light of GS engineering, it is interesting to note that members of the recently described 4,6-α-glucanotransferase (4,6-α-GT) subfamily (which is part of GH70 (Leemhuis *et al.*, 2013a)) do not use sucrose but rather malto-oligosaccharides as the donor substrate, expanding the possibilities for the synthesis of α-glucan type oligosaccharides. Interestingly, models of GTFB (*L. reuteri* 121) and GTFW (*L. reuteri* DSM 20016), generated with Phyre (Kelley & Sternberg, 2009), show that the active sites of these 4,6-α-GT subfamily enzymes are much more open compared to 'true' GSs, likely allowing malto-oligosaccharides to bind as donor substrates in negative subsites (-2, -3, ...)(unpublished results). In order to provide experimental evidence for this hypothesis, we recently initiated crystallization experiments with GTFW-ΔN. Preliminary crystals have already been obtained, but need to be optimized to facilitate X-ray diffraction studies with the crystals, preferably in the presence of malto-oligosaccharides to reveal their binding mode.

Currently, it is only partly understood how glucansucrases synthesize such a large diversity of α-glucans. The polymers differ in the type of glycosidic linkage connecting the glucose moieties (1→2, 1→3, 1→4 or 1→6), and have different degrees of branching, branch length, and molecular mass; moreover often two types of glycosidic linkages are present, and the distribution of branches may be either random or well-defined (*e.g.* in alternan). The structural properties of the α-glucans determine their physicochemical properties, *e.g.* viscosity, stickiness, solubility, mass, etc. Knowledge of the principles underlying GS specificity is therefore essential if we want to be able to use α-glucans in all kinds of applications, especially for engineering GSs to synthesize predefined α-

glucans (*e.g.* for use as prebiotics). The recent advances in GS research have already provided some GS product specificity principles, but many questions remain.

Before GS structures became available, mutation studies attempting to probe and alter GS specificity were based on the proposed similarities with GH13 enzymes. The binding mode(s) of acceptor molecules determines which glycosidic linkage is to be formed during glucosyl transfer; thus, mutations in subsites +1 and +2 likely affect product specificity. Several mutation studies confirmed this principle; for example, the tripeptide following the transition state stabilizing aspartate was identified as a critical region for linkage type specificity, even allowing a shift of linkage type specificity from mainly α(1→4) to mainly α(1→6) (Kralj *et al.*, 2005). The structures of GTF180-ΔN and GTFA-ΔN showed that these residues from motif IV are located near subsite +2, and that the first residue of the tripeptide (S1137 in GTF180-ΔN, N1134 in GTFA-ΔN) interacts with the glucosyl moiety of an acceptor. This residue may therefore be a determinant of linkage type specificity (Kralj *et al.*, 2006; **Chapter 4**). Importantly, the structures also revealed regions (outside the conserved motifs) not targeted before, to be of interest for future mutation studies, such as residues from domain B and from helix α4. Such experiments are currently in progress (X. Meng, personal communication) and will further contribute to our understanding GS product specificity.

A major contribution to GS specificity investigation has been the recent development of techniques to analyze the structures of their initial and final products in great detail. For example, GTFA appears to initially synthesize oligosaccharides with alternating α(1→6) and α(1→4) linkages up to DP12; this pattern is still present in its final product reuteran (Van Leeuwen *et al.*, 2008c, Dobruchowska *et al.*, 2013). A similar alternating pattern is seen for alternansucrase (α(1→6)/α(1→3)). In such cases it is clear that the 'preceding' linkage type strictly determines the 'next' linkage type to be synthesized. In other cases, the linkage type distribution is more random (*e.g.* in GTF180, although its product never contains two consecutive α(1→3) linkages). Also the distribution of branch points may be specific or random; the latter has recently been demonstrated for DSR-E ΔN_{123}-GBD-CD2 (Brison *et al.*, 2013). Together, such studies give valuable insight into the way GSs work, especially during initial stages of α-glucan synthesis.

GS product specificity is also characterized by processivity; the ability to keep binding intermediate products contributes to the final size of the α-glucans. The existence of remote binding sites is a good indication of processivity, but as mentioned before such sites have not been identified yet. Given the size (and heterogeneity) of long oligosaccharides it may be very hard to obtain such information from crystal soaking studies. It is probably more efficient to analyze the products from truncation mutants; *e.g.* deletion of domain V from GTF180-ΔN resulted in a shift from polymer to oligosaccharide formation, or in other words to a less processive enzyme (X. Meng, personal communication).

Crystal structures are now available for GH70 GSs from LAB and from oral *Streptococcus* species. In the oral cavity, GSs likely contribute to the development of caries, especially in the presence of their substrate sucrose, which is present in all kinds of food. In order to prevent caries, one may aim at developing inhibitors that target the oral GSs (Vujičić-Žagar *et al.*, 2010). At the same time, such inhibitors should not inhibit the human salivary and pancreatic GH13 α-amylases. While the active site of α-amylases only bind linear carbohydrates, GSs are able to accommodate branched carbohydrates. A branched carbohydrate may thus be a good starting point for the design of an effective and selective GS inhibitor. The introduction of N-glycosidic linkages instead of the natural O-glycosidic linkages may prevent unwanted degradation.

Taken together, a clearer image of how GSs work most likely can be provided by combining mutagenesis experiments combined with a structural analysis of the reaction products. The design of mutants should be guided by GS crystal structures (also from those synthesizing other classes of α-glucans, and from the 4,6-α-GT subfamily). Site-directed mutagenesis should especially be targeted at acceptor subsites; a detailed structural analysis of the initial and final reaction products is very useful to guide the rational design of such experiments. On the other hand one may apply a directed evolution approach, combined with high-throughput screening methods such as NMR (Irague *et al.*, 2011), to analyze the reaction products. Finally, the application of docking studies with molecular dynamics (MD) to analyze the binding modes of acceptor molecules and intermediate products have already proven to be helpful in explaining the linkage type specificity of GTF180-ΔN (Vujičić-Žagar *et al.*, 2010) and may be applied to other GSs as well.
Ultimately, our knowledge of the GS structure-function relationship may enable to engineer these enzymes such that they synthesize α-glucans (or conjugates thereof) with desired properties. In addition, GS structures are important assets in the development of specific treatments of health-related problems such as caries. The first steps to achieve such goals have been taken, and given the interest for α-glucans especially in prebiotic/probiotic applications, the necessary research efforts are certainly justifiable.

FRUCTANSUCRASES

The CAZy database now lists ~220 putative bacterial fructansucrases (FSs), but only 1 out of 5 have been characterized, while crystal structures are available for only 5 of them. Of the two types of FSs, levansucrases and inulosucrases, most of the research has been done on levansucrases, especially from *Bacillus* species (Anwar *et al.*, 2010a; Van

Hijum *et al.*, 2006). Bacterial FSs are classified in glycoside hydrolase family 68 (GH68), and are related to GH32 enzymes found in plants and fungal species. They typically have molecular masses between 45 and 75 kDa and are constituted of a catalytic domain flanked by N- and C-terminal domains. Research on levan- and inulosucrases from lactic acid bacteria (LAB), synthesizing levan- and inulin-type oligosaccharides and polymers, respectively, started about a decade ago.

The first determined GH68 FS structures were of levansucrases; the structure determination of a truncated inulosucrase, InuJ from *L. johnsonii* NCC 533 (Anwar *et al.*, 2008), so far is the only inulosucrase structure, and the first of an LAB FS. This structure, described in **Chapter 5**, further confirmed the common GH68 three-domain organization with a central catalytic domain having a five-bladed β-propeller topology. As of yet, the functions of the N- and C-terminal domains, which both were truncated in the crystallized construct, remain unknown. Crystallization experiments with constructs including the full N-terminal domain have been unsuccessful so far; the structure determination of a full-length LAB FS remains an interesting target.

To gain insight into the substrate specificity of an inulosucrase, we soaked crystals of a catalytically inactive InuJ mutant with the substrate sucrose. The structure of the enzyme-substrate complex revealed that shared substrate specificity between levan- and inulosucrases is fully explained by the virtually identical architecture of the active site. The deep active site pocket, at the bottom of a wide funnel, is tailored to bind the substrate sucrose in subsites -1 and +1 (with the fructosyl moiety bound in subsite -1). The structure also further confirms the double-displacement reaction mechanism that had been proposed for GH68 FSs before (Meng & Fütterer, 2003; Van Hijum *et al.*, 2006). This two-step mechanism involves three absolutely conserved residues (Asp/Glu/Asp) functioning as nucleophile, general acid/base, and transition state stabilizing residue, respectively. Cleavage of the substrate results in a covalent fructosyl-enzyme intermediate; the fructosyl moiety is then transferred to an incoming acceptor molecule (occupying subsites +1, +2, …), *via* a β(2→1) or β(2→6) linkage. In this way, the product is elongated by one fructosyl unit per reaction cycle (Ozimek *et al.*, 2006b). Alternatively, the covalent intermediate can be hydrolyzed, releasing fructose.

Understanding how different FSs synthesize β-fructans of different size, linkage type (β(2→1) or β(2→6)) and branching distribution, would enable the engineering of FSs to synthesize β-fructans with desired properties. With crystal structures becoming available for GH68 and the related GH32 enzymes during the last decade, we are getting more insight into this matter.

Regarding product size specificity, the architecture of the active site and its accessibility for water determine the transfructosylation/hydrolysis ratio and therefore the size of the final product. Initially, researchers focused on residues interacting with sugar moieties in subsites -1 and +1. For example, an almost strictly conserved arginine residue (R360 in levansucrase SacB from *B. subtilis*, GH68), located near subsite +1, was attributed a key role for transfructosylation (Meng & Fütterer, 2008). This residue has been the subject in many mutation studies, showing that its mutation affected product size. Recently, a glutamine at the equivalent position in GH32 enzymes, together with a nearby aspartate, was proposed to be of key importance for the product specificity in 6-SST/6-SFT from *Pachysandra terminalis* (Lammens *et al.*, 2012) and in β-fructofuranosidase from *Schwanniomyces occidentalis* (Álvaro-Benito *et al.*, 2012). However, more remote sugar binding sites likely also play a role, as they may increase the affinity for acceptors and intermediate products. Several mutation studies support such a concept; our sucrose-soaked InuJ crystals provide further evidence, as a second sucrose molecule was observed near residues from a non-conserved loop (1B-1C) in InuJ. The presence of remote acceptor binding sites may enable FS enzymes to be processive and synthesize large β-fructan polymers. Unfortunately, structures of complexes of FSs with long oligosaccharides (DP >2) have not been reported yet; the flexibility (and heterogeneity) of such compounds may well prevent successful crystallization.

To gain insight in linkage type specificity, we analyzed the structure of InuJ complexed with the inulin-type trisaccharide 1-kestose, the product of the first transfructosylation step catalyzed by InuJ, occupying subsites -1, +1 and +2. Although not representing a genuine acceptor binding mode (then the trisaccharide should occupy subsites +1 to +3), it does show how a fructosyl moiety in subsite +1 would be oriented after β(2-1) linkage formation. Surprisingly, the residues making hydrogen bond interactions with the trisaccharide are conserved between GH68 inulo- and levansucrases. We propose that other structural elements, such as the previously mentioned 1B-1C loop, are important for guiding acceptor molecules in the right orientation for transfructosylation. Mutation of residues in this loop indeed decreased the transfructosylation activity of InuJ. This concept of remote binding sites is also supported by studies on the related GH32 fructosyltransferases, where residues from surface loops at the rim of the active site funnel are near sugar moieties occupying higher subsites (+2, +3) and seem to contribute to specificity (Chuankhayan *et al.*, 2010; Lammens *et al.*, 2012). Interestingly, many levansucrases, while synthesizing levan, also synthesize short inulin-type oligosaccharides; this observation complicates our understanding of linkage type specificity in FSs. Thus, although some residues have been attributed a key role for transfructosylation, so far it has not been possible to pinpoint residues that determine whether fructosyl transfer occurs with a β(2→1) or β(2→6) linkage. To this end, it would

be most interesting to determine the structure of a GH68 levansucrase in complex with a levan-type oligosaccharide (*e.g.* 6-kestose). The levansucrases from *L. reuteri* 121 (Ozimek *et al.*, 2006b) or from *L. gasseri* DSM 20077 (Anwar *et al.*, 2010b) are good targets for initiating such studies. Finally, it may not only be the enzyme architecture, but also the structural properties of (intermediate) products that play a role in product linkage specificity; in this light, differences in flexibility and helical preference between levans and inulins are of interest. In addition to determining crystal structures, our understanding of FS linkage type specificity may be further advanced by the application of docking / molecular dynamics studies with GH68 FSs and their substrates.
Another approach to study the product specificity of FSs is to analyze the products of mutants, guided by known crystal structures. Anwar *et al.* (2012) generated a set of mutants of *L. reuteri* 121 inulosucrase, focusing on residues that were non-conserved with levansucrase from the same bacterial species, and based on the structure of InuJ. This study revealed that it was easy to obtain variants with higher transfructosylation specificity, higher catalytic rate and different product sizes, while leaving the linkage type (β(2→1) unaltered. Still, it remains to be discovered what makes an FS to be either an inulosucrase or a levansucrase.

Together, by combining structural and mutation studies of LAB FSs with a detailed analysis of their products, it should be possible to engineer FS enzymes such that they synthesize products with desired properties, from a cheap substrate (sucrose). Like the α-glucans synthesized by glucansucrases, such (modified) fructans hold great potential especially for prebiotics-related applications.